Centrality in networks of urban streets

Paolo Crucitti, Vito Latora, and Sergio Porta
Scuola Superiore di Catania, Italy Dipartimento di Fisica e Astronomia, Università di Catania, and INFN Sezione di Catania, Italy and Dipartimento di Progettazione dell'Architettura, Politecnico di Milano, Italy

(Received 6 September 2005; accepted 11 November 2005; published online 31 March 2006)

Centrality has revealed crucial for understanding the structural properties of complex relational networks. Centrality is also relevant for various spatial factors affecting human life and behaviors in cities. Here, we present a comprehensive study of centrality distributions over geographic networks of urban streets. Five different measures of centrality, namely degree, closeness, betweenness, straightness and information, are compared over 18 1-square-mile samples of different world cities. Samples are represented by primal geographic graphs, i.e., valued graphs defined by metric rather than topologic distance where intersections are turned into nodes and streets into edges. The spatial behavior of centrality indices over the networks is investigated graphically by means of color-coded maps. The results indicate that a spatial analysis, that we term multiple centrality assessment, grounded not on a single but on a set of different centrality indices, allows an extended comprehension of the city structure, nicely capturing the skeleton of most central routes and subareas that so much impacts on spatial cognition and on collective dynamical behaviors. Statistically, closeness, straightness and betweenness turn out to follow similar functional distribution in all cases, despite the extreme diversity of the considered cities. Conversely, information is found to be exponential in planned cities and to follow a power-law scaling in self-organized cities. Hierarchical clustering analysis, based either on the Gini coefficients of the centrality distributions, or on the correlation between different centrality measures, is able to characterize classes of cities.
 [DOI: 10.1063/1.2150162]

Centrality measures serve to quantify that in a network some nodes are more important (central) than others. The idea of centrality was first introduced in the context of social systems, where it was assumed a relation between the location of an individual in the network and its influence and/or power in group processes. Since then, various centrality measures have been proposed over the years to quantify the structural centrality of an individual in a social network, and the issue of centrality has found many applications also in biology and technology. When dealing with urban street patterns, centrality has been investigated in relational (topological) networks only, neglecting a fundamental aspect of the system as the geography. Here, we consider urban street patterns as spatial networks, i.e., networks embedded in the real space (whose nodes occupy a precise position in a two-dimensional Euclidean space, and whose edges are real physical connections). In such an approach, a city is transformed into a spatial graph by mapping the intersections into the graph nodes and the roads into links between nodes. By using a set of different centrality indices (multiple centrality assessment), extended or defined on purpose for spatial graphs, it is possible to spot the relevant places of a city. Relevant places means, places closer to other places (closeness centrality), places that are structurally made to be traversed (betweenness centrality), places whose route to other places deviates less from the virtual straight route (straightness centrality), and places whose deactivation affects the structural properties of the system (information centrality). Moreover, by investigating how centrality is distributed among the nodes of the graph, and how the different centrality indices are correlated, it is possible to characterize classes of cities. In particular, we have found large differences between self-organized and single-planned cities in the distribution of information centrality. The centrality analysis hereby presented opens up to the in depth investigation of the correlation between the structural properties of the network and the relevant dynamics on the network like pedestrian and/or vehicular flows, retail commerce vitality, land-use separation or urban crime. We expect that some of these factors are more strictly correlated to some centrality indices than to others, thus giving informed indications on the actions that can be performed in order to increase the desired factors, as economic development, and to hinder the undesired ones, as crime rate.

I. INTRODUCTION

The science of networks has been witnessing a rapid development in recent years: the metaphor of the network, with all the power of its mathematical devices, has been applied to complex, self-organized systems as diverse as social, biological, technological and economic, leading to the achievement of several unexpected results.[1–3] In particular, the issue of centrality in networks has remained pivotal, since its introduction in a part of the studies of humanities named structural sociology.[4] The idea of centrality was first applied to human communication by Bavelas[5] who was interested in the characterization of the communication in small groups of people and assumed a relation between structural centrality and influence and/or power in group processes. Since then, various measures of structural centrality have been proposed over the years to quantify the importance of an individual in a social network.[4] Currently, centrality is a fundamental concept in network analysis though with a different purpose: while in the past the role and identity of central nodes were investigated, now the emphasis is more shifted to the distribution of centrality values through all nodes. Centrality, as such, is treated like a shared resource of the network community, like wealth in nations, with the focus being on the homogeneity and/or heterogeneity of distributions.[1] In urban planning and design, as well as in economic geography, centrality, though under different terms like accessibility, transport cost or effort, has entered the scene stressing the idea that some places are more important than others because they are more central;[6] all these approaches have been following a primal representation of spatial systems, where punctual geographic entities (street intersections, settlements) are turned into nodes and their linear connections (streets, infrastructures) into edges. A pioneering discussion of centrality as inherent to urban design in the analysis of spatial systems has been successfully operated after Hillier and Hanson seminal work on cities[7] since the mid-1980s. *Space Syntax*, the related methodology of urban analysis, has been raising growing evidence of the correlation between the so-called integration of urban spaces, a closeness centrality in all respects, and phenomena as diverse as crime rates, pedestrian and vehicular flows, retail commerce vitality and human way-finding capacity.[8] The Space Syntax approach follows a dual representation of street networks where streets are turned into nodes and intersections into edges. An outcome of the dual nature of Space Syntax is that the node degree is not limited by physical constraints, since one street has a conceptually unlimited number of intersections; this property makes it possible to witness the emerging of power laws in degree distributions[9–11] that have been found to be a distinct feature of other nongeographic systems.[1–3,12] On the other hand, the dual character leads Space Syntax to the abandonment of metric distance, a street is one node no matter its real length. Metric distance, conversely, was the core of most if not all territorial studies[13] and is a key ingredient of spatial networks.[14] In this paper we propose a primal network analysis of urban street systems within a properly geographic framework based on metric distances. In the primal representation urban street patterns are turned into undirected, valued, primal graphs, where intersections are nodes and streets are edges. We show that by using a set of various centrality measures, it is possible to characterize and discuss urban networks within the same framework of all other complex systems of a nongeographic nature.

II. THE NETWORK APPROACH: SPACE SYNTAX (SS)

The network approach has been broadly used in urban studies. Since the early 1960s, a lot of research has been spent trying to link the allocation of land uses to population growth through lines of transportation,[6] or seeking the prediction of transportation flows given several topological and geometric characteristics of traffic channels[15] or eventually investigating the exchanges of goods and habits between settlements in the geographic space even in historical eras.[16]

Urban design as a discipline, beside some few theoretical efforts[17–19] has not contributed that much to the picture in direct operational terms, with one quite relevant exception. In fact, after the seminal work of Hillier and Hanson,[7] a rather consistent application of the network approach to cities, neighborhoods, streets and even single buildings, has been developed under the notion of Space Syntax (SS).

The network analysis, when applied to territorial cases, has mostly followed a primal graph representation, where intersections (or settlements) are turned into the nodes of a graph and streets (or relationships) into edges. That representation seems to be the most intuitive for networks characterized by a strong connection to the geographic dimensions, which is to say networks where distance must be measured not just in topological terms (steps), like, for instance in social systems, but rather in properly spatial terms (meters, miles), like in urban street systems. It might appear paradoxical, though, that Space Syntax, the flagship application of urban design to the network analysis of city spaces, did follow the opposite direction, being based on a dual graph representation of urban street patterns. In this representation, axial lines that represent generalized streets (more exactly, “lines of sight” or “lines of unobstructed movement” along mapped streets) are turned into nodes, and intersections between each pair of axial lines into edges. More precisely, Space Syntax is based on the four steps illustrated in Fig. 1, top panel.

(i) Based on the idea that the basic spatial unit is the line of sight (or unobstructed movement),[11] the street pattern (1) is transformed into an axial map (2), still a primal representation though not properly a graph.

(ii) The axial map (2) is transformed into a dual graph (3), called the connectivity graph. This is an undirected graph made by N nodes, the number of axial lines, and K links representing the intersections between couples of axial lines.

(iii) Though not limited to just one index, the core of the Space Syntax methodology, when applied to street networks, is the index of integration, which is stated to be “so fundamental that it is probably in itself the key to most aspects of human spatial organization.”[8] The global integration INT_i of street (axial line) i is defined as[20,21]

$$\mathrm{INT}_i = \frac{D_N}{RA_i}, \tag{1}$$

where D_N is a normalizion factor depending solely on N, and RA_i is the so-called relative asymmetry. This latter quantity is defined as

$$RA_i = \frac{2(MD_i - 1)}{N-2}, \quad MD_i = \frac{1}{N-1}\sum_{j=1}^{N} s_{ij}, \tag{2}$$

where s_{ij} represents the length of the shortest journey route between two streets in the city, and is given by

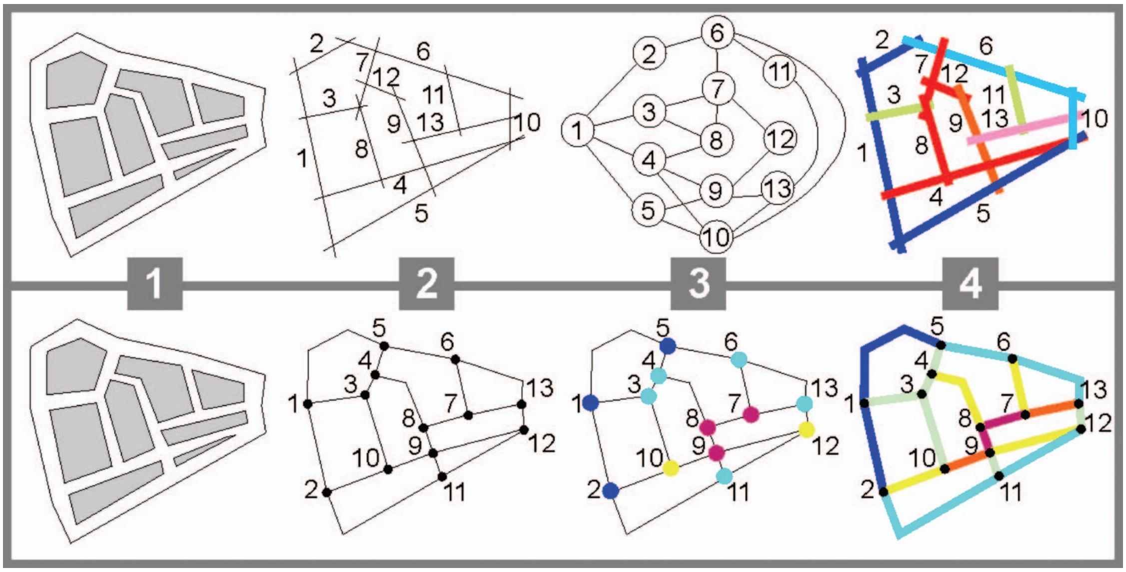

FIG. 1. (Color) Sketch of the basic steps in the space syntax (SS) approach (top panel) and in the proposed multiple centrality assessment (MCA) (bottom panel). In this latter approach, node centrality values are calculated directly on the primal graph representation (2), avoiding the dual passage. Color/coded values are reported on the nodes of the primal graph (3). Centrality values can then be reported on nodes (3) or on edges (4). See text for details.

the smallest number of steps (i.e., the fewest changes in direction) between the correspective nodes of the dual graph (3). In practice, the integration index, as defined in Eq. (1), turns out to be nothing else than a differently normalized version of the closeness centrality [Eq. (4)] that we will discuss in the next section in the context of the multiple centrality assessment (MCA).

(iv) The integration of each node in the connectivity graph (3) is calculated, and color/coded values are reported back on the axial map (4), giving rise to the final primal-like color/coded representation.

An example of the axial map for the city of Cairo is reported in Fig. 2. Here, the streets in red represent the most central nodes of the dual graph in terms of the global integration defined in Eq. (1), while those in blue are the less central.

In the past 20 years, Space Syntax has been applied to many urban cases establishing a significant correlation between the topological centrality of streets and phenomena as diverse as their popularity (pedestrian and vehicular flows), human way-finding, safety against microcriminality, retail commerce vitality, activity separation and pollution.[22] Shortcomings as well as benefits of this approach have been often remarked.[11,20,23–27] The main problem with SS is that it does not account for metric distances. One street is turned in one node (a dimensionless entity). Only the topology of the system is considered, relations between nodes are just step distances, which leads to a substantial underestimation of the performative motivations of collective behaviors (deeply affected by the metric factor) in favor of their sole cognitive motivations (more affected by the pure relational factor).[26] Another problem with the SS is that the outcomes are mostly based on a single index, that of integration-closeness centrality; this index turns out to be so vulnerable to the edge effect—the distortion that gathers higher centrality values around the geometrical center of the image—that it would make the whole analysis meaningless without the implementation of a generalization process, which is what axial mapping does. On the other hand, having implemented the axial mapping process, the emergence of central routes in SS is not the outcome of the natural flow of centrality; rather it is heavily affected by the axial mapping rationale, in particular, it is impossible to account for variations along the same line,

FIG. 2. (Color) An application of the SS. The axial map for the city of Cairo. Streets in red represent central nodes in the dual graph according to the integration index of Eq. (1). Space Syntax measures have been calculated using the Axwoman 1.0 extension of Arcview 3.1 over a handmade axial map of the 1-square-mile map of central Cairo. Axwoman is a software developed in the Center for Advanced Spatial Analysis (CASA), University College, London.

while the line does intersect other lines. A clear example is the axial line number 4 in the top panel of Fig. 1, this route to be a single unit, it is an outcome of axial mapping, and not of the centrality flow over the system.

Finally, the SS dual approach is fundamentally different from the traditional network representation of geographic systems, which is primal: an immense amount of information is currently available in this format which can directly support the MCA analysis.

III. THE MULTIPLE CENTRALITY ASSESSMENT

The multiple centrality assessment relies on three basic principles[26,27] as follows:

(1) primal graphs, rather than dual;
(2) metric distance, rather than topologic;
(3) many centrality indices, rather than mainly closeness.

The method is based on the four steps illustrated in Fig. 1, bottom panel.

(i) The urban street pattern (1) is trasformed into an undirected valued primal graph (2): the intersections are turned into graph nodes, and streets are the edges. Edges follow the footprints of real streets. The length of each street is associated to the corresponding edge. In this way both the topology and the geography (metric distances) of the system are considered.
(ii) The evaluation of the importance of a node is based on different node centrality measures, namely closeness C^C, betweenness C^B, straightness C^S and information C^I, discussed below. The four centrality measures are calculated for each node of the primal graph (2).
(iii) Color/coded values are reported on the nodes of the primal graph (2), giving rise to four figures as that reported in (3), one for each of the four centrality measures. Of these nature are, for instance, Fig. 4 and Fig. 5.
(iv) The final layout can either map node (3), as well as edge centrality (4). In this latter case, the centrality of an edge is calculated as the average of its couple of endnodes; this simple procedure highlights a deep character of spatial networks when represented in such a primal way: one edge exchanges with the system only at nodes, so its relational properties as a system's component entirely depends on its endnodes' importance.

The following is a list of the centrality measures we have adopted. The definitions are given in terms of an undirected, valued (weighted) graph G, of N nodes and K edges. The graph is described by the adjacency $N\times N$ matrix A, whose entry a_{ij} is equal to 1 when there is an edge between i and j and 0 otherwise, and by a $N\times N$ matrix L, whose entry l_{ij} is the value associated to the edge, in our case the metric length of the street connecting i and j.

Degree centrality, C^D, is the simplest definition of node centrality. It is based on the idea that important nodes have the largest number of ties to other nodes in the graph. The degree centrality of i is defined as:[4,28,29]

$$C_i^D = \frac{\sum_{j=1,N} a_{ij}}{N-1} = \frac{k_i}{N-1}, \tag{3}$$

where k_i is the degree of node i, i.e., the number of nodes adjacent to i. Degree centrality is not particularly relevant in primal urban networks where node degrees are limited by geographic constraints.

Closeness centrality, C^C, measures to which extent a node i is near to all the other nodes along the shortest paths, and is defined as[4,30]

$$C_i^C = \frac{N-1}{\sum_{j\in G, j\neq i} d_{ij}}, \tag{4}$$

where d_{ij} is the shortest path length between i and j, defined, in a valued graph, as the smallest sum of the edges length throughout all the possible paths in the graph between i and j.

Betweenness centrality, C^B, is based on the idea that a node is central if it lies between many other nodes, in the sense that it is traversed by many of the shortest paths connecting couples of nodes. The betweenness centrality of node i is[29]

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j,k\in G, j\neq k\neq i} n_{jk}(i)/n_{jk}, \tag{5}$$

where n_{jk} is the number of shortest paths between j and k, and $n_{jk}(i)$ is the number of shortest paths between j and k that contain node i.

Straightness centrality, C^S, originates from the idea that the efficiency in the communication between two nodes i and j is equal to the inverse of the shortest path length d_{ij}.[31] The straightness centrality of node i is defined as

$$C_i^S = \frac{1}{N-1} \sum_{j\in G, j\neq i} d_{ij}^{\mathrm{Eucl}}/d_{ij}, \tag{6}$$

where d_{ij}^{Eucl} is the Euclidean distance between nodes i and j along a straight line, and we have adopted a normalization recently proposed for geographic networks.[32] This measure captures to which extent the connecting route between nodes i and j deviates from the virtual straight route.

Information centrality, C^I, is a measure introduced in Ref. 33, and relating a node importance to the ability of the network to respond to the deactivation of the node. The network performance, before and after a certain node is deactivated, is measured by the efficiency of the graph G.[31,34] The information centrality of node i is defined as the relative drop in the network efficiency caused by the removal from G of the edges incident in i,

$$C_i^I = \frac{\Delta E}{E} = \frac{E[G]-E[G']}{E[G]}, \tag{7}$$

where the efficiency of a graph G is defined as

TABLE I. Basic properties of the primal graphs obtained from the 18 1-square-mile samples of the different world cities considered. N is the number of nodes, K is the number of edges, $<l$ and σ_l are, respectively, average edge length and standard deviation.

	Case	N	K	$\langle l \rangle$	σ_l
1	Ahmedabad	2870	4387	27.59	17.03
2	Barcelona	210	323	112.01	50.16
3	Bologna	541	773	66.26	39.12
4	Brasilia	179	230	134.39	90.02
5	Cairo	1496	2255	37.47	25.42
6	Los Angeles	240	340	113.87	64.86
7	London	488	730	72.33	39.52
8	New Delhi	252	334	96.56	75.73
9	New York	248	419	86.33	104.11
10	Paris	335	494	89.29	56.06
11	Richmond	697	1086	57.65	45.49
12	Savannah	584	958	64.77	39.49
13	Seoul	869	1307	52.12	29.57
14	San Francisco	169	271	140.91	119.93
15	Venice	1840	2407	31.25	23.58
16	Vienna	467	692	72.16	39.71
17	Washington	192	303	119.94	53.75
18	Walnut Creek	169	197	127.57	86.01

$$E[G] = \frac{1}{N(N-1)} \sum_{i,j \in G, i \neq j} d_{ij}^{\mathrm{Eucl}} / d_{ij} \tag{8}$$

and where G' is the graph with N nodes and $K-k_i$ edges obtained by removing from the original graph G the edges incident in node i. An advantange of using the efficiency to measure the performance of a graph is that $E[G]$ is finite even for disconnected graphs.

IV. APPLICATION TO 1-SQUARE MILE MAPS

We have selected 18 1-square-mile samples of different world cities from the book by Jacobs,[35] imported them in a GIS (Geographic Information System) environment and constructed primal graphs of street networks using a road-centerline-between-nodes format.[36] The studied cities are listed in Table I together with the basic properties of the derived graphs, number of nodes N, links K, average edge (street) length $\langle l \rangle$, and standard deviation σ_l. The cases considered exhibit striking differences in terms of cultural, social, economic, religious, and geographic context. In particular, they can be roughly divided into two large classes, (1) patterns grown throughout a largely self-organized, fine-grained historical process, out of the control of any central agency; (2) patterns realized over a short period of time as the result of a single plan, and usually exhibiting a regular gridlike, structure. Ahmedabad, Cairo, and Venice are the most representative examples of self-organized patterns, while Los Angeles, Richmond, and San Francisco are typical examples of mostly planned patterns. The basic characteristics of the derived graphs, N, K, $\langle l \rangle$, σ_l assume widely different values, notwithstanding the fact we have considered the same amount of land. In Fig. 3 we report the edges length distribution $P(l)$ for the two different classes of cities. In particular we show the cases of Ahmedabad and Cairo as

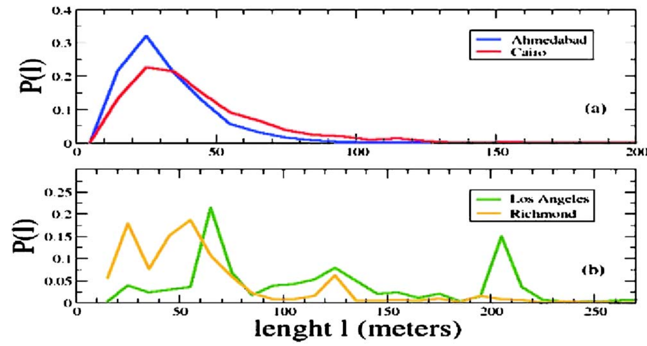

FIG. 3. Length distributions for (a) two self-organized cities (Ahmedabad and Cairo), and (b) two planned cities (Los Angeles and Richmond). Length distributions $P(l)$ are defined by $N(l)/N$, where $N(l)$ is the number of edges whose length is in the range ($l-5$ meters; $l+5$ meters).

self-organized cities, and Los Angeles and Richmond as mostly planned cities. Cities of the first class show single peak distributions, while cities of the second one show a multimodal distribution, due to their grid pattern. Other details on the structural properties of such networks can be found in Refs. 37 and 38. Finally, for each of the 18 cities, we have evaluated five node centrality indices, namely C^D, C^C, C^B, C^S, and C^I. As already mentioned, degree centrality turns out to be not particularly relevant in primal urban networks where node degrees are limited by geographic constraints. For this reason, in the following we will mainly focus on four indices, C^C, C^B, C^S, and C^I.

A. The spatial distribution of centralities

The spatial distributions of node centralities can be graphically illustrated by means of GIS supported color-coded maps, in which one of eight different colors is plotted on each node of the graph. In Figs. 4 and 5 are shown, respectively, the case of Cairo and Richmond. The figures for the remaining cities can be downloaded from our website.[39] The colors represent eight classes of nodes with different values of the centrality index C. The classes, defined in terms of multiples of the standard deviations σ from the average, are $[-\infty,-3\sigma]$, $[-3\sigma,-2\sigma]$, $[-2\sigma,-\sigma]$, $[-\sigma,0]$, $[0,\sigma]$, $[\sigma,2\sigma]$, $[2\sigma,3\sigma]$, $[3\sigma,\infty]$, and the corresponding colors are reported in the figure legends. In both cases, C^C exhibits a strong trend to group higher scores at the center of the image [Fig. 4 panel (a) and Fig. 5 panel (a)]. This is due to the artificial boundaries imposed by the 1-square-mile maps representation and to the same nature of the closeness centrality. Edge effects are also present, although less relevant, in the other centrality measures (see for instance the contour nodes in Fig. 5, panel (b), (c), and (d).) The spatial distribution of C^B nicely captures the continuity of prominent urban routes across a number of intersections, changes in direction and focal urban spots. This is visible both in Cairo, Fig. 4(b), and in Richmond, Fig. 5(b). In particular, in Richmond C^B clearly identifies the primary structure of movement channels as different to that of secondary, local routes. The same happens in Ahmedabad and Seoul. Among the other cities not shown, C^B is particularly effective in Venice, where most popular walking paths and squares ("campi"), and the Rialto

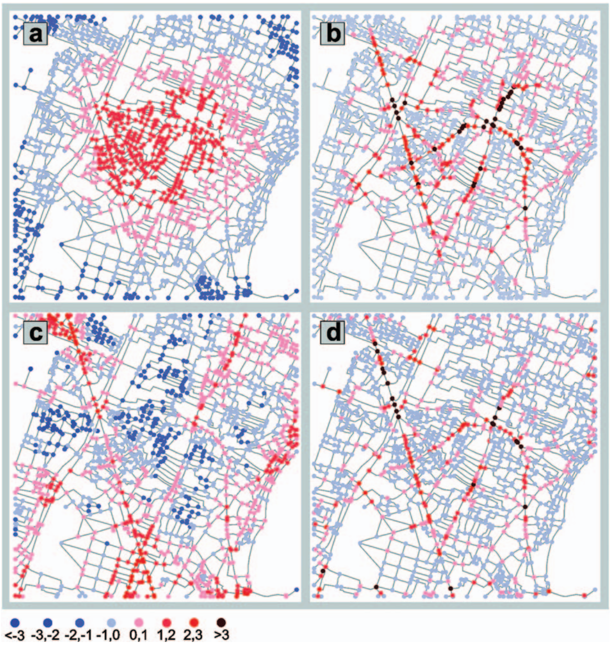

FIG. 4. (Color) Thematic color map representing the spatial distributions of centrality in Cairo, an example of a largely self-organized city. The four indices of node centrality, (a) closeness C^C, (b) betweenness C^B, (c) straightness C^S, and (d) information C^I, used in the MCA, are visually compared over the primal graph. Different colors represent classes of nodes with different values of the centrality index. The classes are defined in terms of multiples of standard deviations from the average, as reported in the color legend.

bridge over the Canal Grande, emerge along the red nodes routes. The spatial distribution of C^S depicts both linear routes and focal areas in the urban system [Figs. 4(c) and 5(c)], C^S takes high values along the main axes, even higher at their intersections. Finally C^I, although based on a different concept of centrality, exhibits a spatial distribution that is in many cases similar to that of C^B. This is especially evident in Cairo Fig. 4(d), as well as in Ahmedabad and Venice. Notwithstanding the similarities in the color maps, the two measures exhibit radically different statistical distributions.

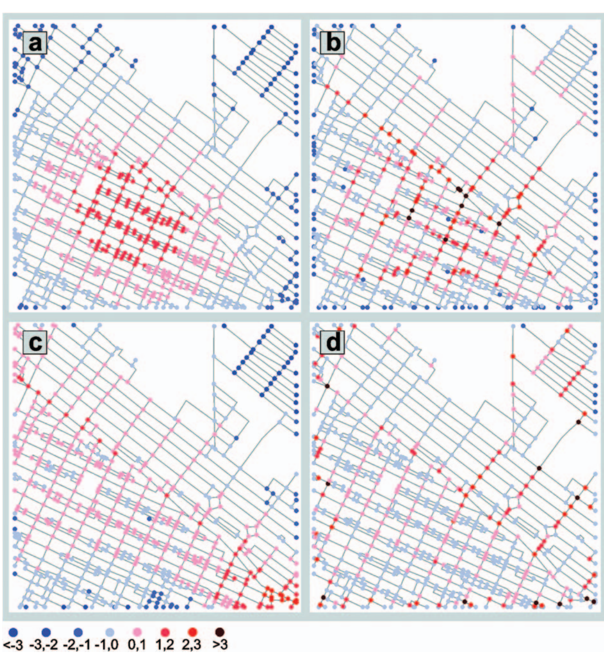

FIG. 5. (Color) Thematic color map representing the spatial distributions of the same four centrality indices of Fig. 4 in Richmond, an example of a mostly planned city.

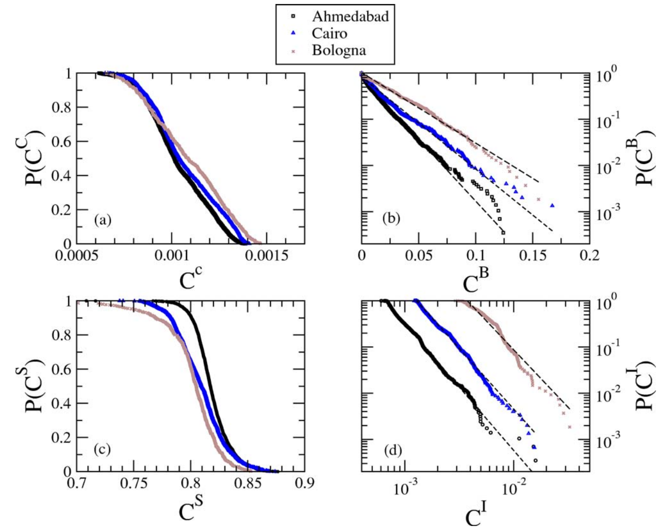

FIG. 6. Cumulative distributions of (a) closeness C^C, (b) betweenness C^B, (c) straightness C^S, and (d) information C^I for three planned cities, Ahmedabad, Cairo, and Bologna. The dashed lines in panels (b) are exponential fits to the betweenness distributions, while the dashed lines in panel (d) are power law fits to the information centrality.

B. The statistical distribution of centralities

In Figs. 6 and 7 we report an example of the cumulative distributions of centrality obtained for the two categories of cities, self-organized cities (Ahmedabad, Cairo, and Bolo-

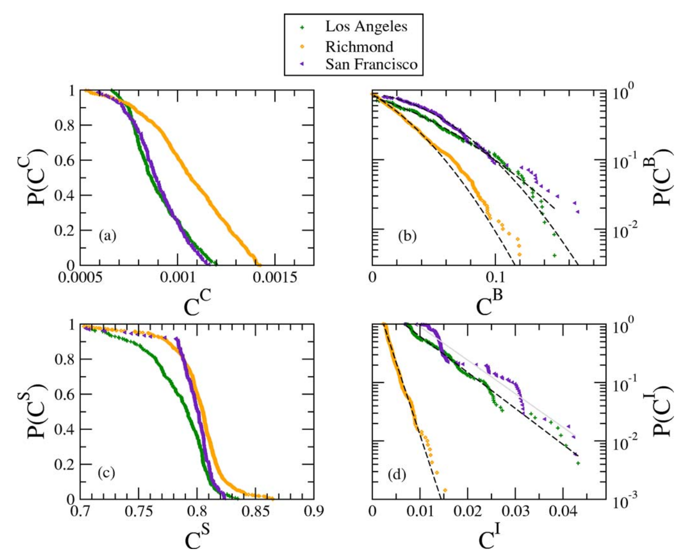

FIG. 7. Cumulative distributions of (a) closeness C^C, (b) betweenness C^B, (c) straightness C^S, and (d) information C^I for three planned cities, Los Angeles, Richmond, and San Francisco. The dashed lines in panels (b) are Gaussian fits to the betweenness distributions, while the dashed lines in panel (d) are exponential fits to the information centrality.

TABLE II. Results of the fitting to the betweenness and information centrality distributions for the most representative examples of self-organized (Ahmedabad, Bologna, Cairo, Venice) and single-planned patterns (Los Angeles, Richmond, New York, and San Francisco). See discussion in the text.

Case	C^B	C^I
Ahmedabad	$s=0.016$	$\gamma=2.74$
Bologna	$s=0.028$	$\gamma=2.49$
Cairo	$s=0.022$	$\gamma=2.63$
Venice	$s=0.044$	$\gamma=1.49$
Los Angeles	$\sigma=0.078$	$s=0.007$
Richmond	$\sigma=0.049$	$s=0.002$
New York	$\sigma=0.052$	$s=0.003$
San Francisco	$\sigma=0.068$	$s=0.008$

gna) and single-planned cities (Los Angeles, Richmond, and San Francisco). The cumulative distribution $P(C)$ is defined as

$$P(C) = \int_C^{+\infty} \frac{N(C')}{N} dC' , \tag{9}$$

where $N(C)$ is the number of nodes with centrality equal to C. Closeness, straightness, and betweenness distributions are quite similar in both self-organized and planned cities, despite the diversity of the two cases in sociocultural and economic terms could not be deeper. In particular, C^B exhibits a single scale distribution[40] both in self-organized and planned cities, the former having an exponential distribution, the latter having a Gaussian distribution. The dashed lines in Fig. 6 panel (b) are exponential fits to the empirical distributions of the form $P(C) \sim \exp(-C/s)$, with coefficient s, respectively, equal to $s_{\mathrm{Ahm}}=0.016$, $s_{\mathrm{Cai}}=0.022$, $s_{\mathrm{Bol}}=0.028$. The dashed lines in Fig. 7 panel (b) are Gaussian fits, $P(C) \sim \exp(-1/2x^2/\sigma^2)$, with $\sigma_{\mathrm{LA}}=0.078$, $\sigma_{\mathrm{Rich}}=0.049$, $\sigma_{\mathrm{SF}}=0.068$. On the other hand, the information centrality distributions notably differentiate self-organized cities from planned ones, being broad-scale (power law) in the first case, and single-scale (exponential) in the second case. The dashed lines in the log-log plot of Fig. 6 panel (d) are power law fits $P(C) \sim C^{-\gamma}$ to the empirical distributions of self-organized cities with exponents: $\gamma_{\mathrm{Ahm}}=2.74$, $\gamma_{\mathrm{Cai}}=2.63$, $\gamma_{\mathrm{Bol}}=2.49$, whereas the dashed lines in Fig. 7 panel (d) are exponential fits of of the form $P(C) \sim \exp(-C/s)$ with coefficients, $s_{\mathrm{LA}}=0.007$, $s_{\mathrm{Rich}}=0.002$, $s_{\mathrm{SF}}=0.008$. Interestingly enough the results we have found are also valid for a self-organized city with strong environmental and physical contraints as Venice. In fact, notwithstanding the peculiar historic process by which the city grew up by the colonization of hundreds of small islands around a major canal, the distribution of betweenness in Venice is exponential, while $P(C^I)$ is a power law with the smallest value of the exponent found, namely $\gamma_{\mathrm{Ven}}=1.49$. The results of the fittings are reported in Table II.

Similar results have been obtained by modelling planned cities as regular triangular, square or rectangular lattices, and self-organized cities as disordered[27] or growing networks.[17,41,42] The identified power laws in the information centrality, similar to those found in the degree[1,12] and in the

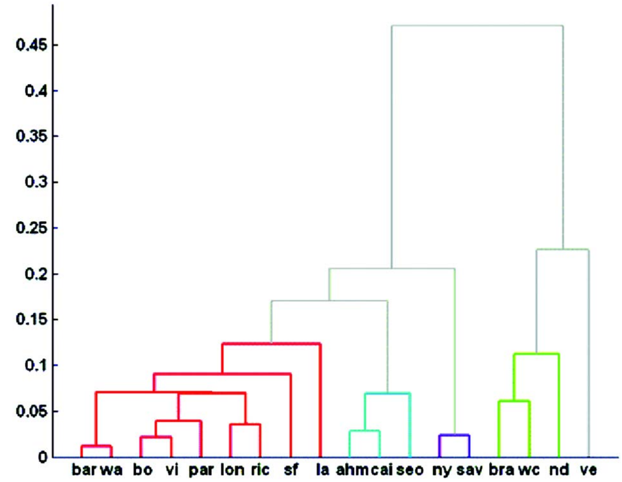

FIG. 8. Hierarchical tree (dendrogram) based on the four Gini coefficients of the centrality distributions. The complete linkage method, based on the largest distance between objects in different clusters, has been applied. Choosing a maximum distance equal to 0.15 for two cities to belong to the same cluster one obtains five different clusters. See discussion in the text.

betweenness[43] distributions of some nonspatial graphs, indicate a highly uneven distribution of the centrality over self-organized networks: most nodes have low centrality scores and coexist with a few nodes with high values of C^I.

Inequalities in the distribution of the four centrality indices among the nodes of the network can be quantified consistently by evaluating the Gini coefficients of the distributions.[44] The Gini coefficient, g, is an index commonly adopted to measure inequalities of a given resource among the individuals of a population. It can be calculated by comparing the Lorenz curve of a ranked empirical distribution, i.e., a curve that shows, for the bottom $x\%$ of individuals, the percentage $y\%$ of the total resource which they have, with the line of perfect equality.[44] The coefficient g ranges from a minimum value of zero, when all individuals are equal, to a maximum value of 1, in a population in which every individual except one has a size of zero. For each city we have calculated four Gini coefficients, g^C, g^B, g^S, g^I, one for each of the centrality distributions.

For example, in the case of the information centrality, we have obtained a Gini coefficient g^I equal to 0.12 for New York, 0.19 for Richmond, and 0.23 for Cairo, thus indicating that Cairo shows a distribution more heterogeneous than those of Richmond and New York. In Fig. 8 we show the results of a hierarchical clustering analysis in which the distance between two cities m and n, D_{mn}, is defined in terms of the four Gini coefficients as

$$D_{mn} = \sqrt{\sum_{i=1}^{4} \frac{(g_m^i - g_n^i)^2}{\alpha_i}}. \tag{10}$$

where g_m^i, $i=1,2,3,4$ are, respectively, the four Gini coefficients $g_m^C, g_m^B, g_m^S, g_m^I$ for the city mth, and α_i, $i=1,2,3,4$ are four normalization constants. The results reported in Fig. 8 have been obtained by setting the four constants all equal to 1. We have also investigated a different normalization in which each α_i is equal to the maximum (with respect to m and n) of $(g_m^i - g_n^i)^2$. We have used a complete linkage

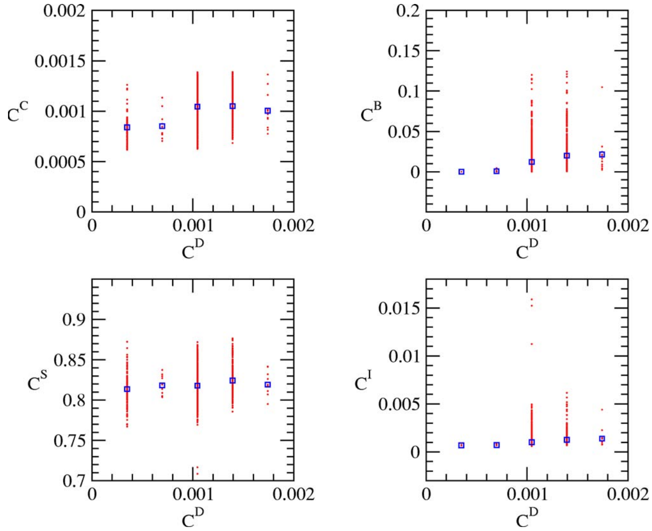

FIG. 9. Scatter plots representing the correlations between degree centrality, C^D, and the four other measures of centrality used in the MCA, namely C^C, C^B, C^S, C^I. The results reported are for the city of Ahmedabad. The squares are averages over nodes with the same degree.

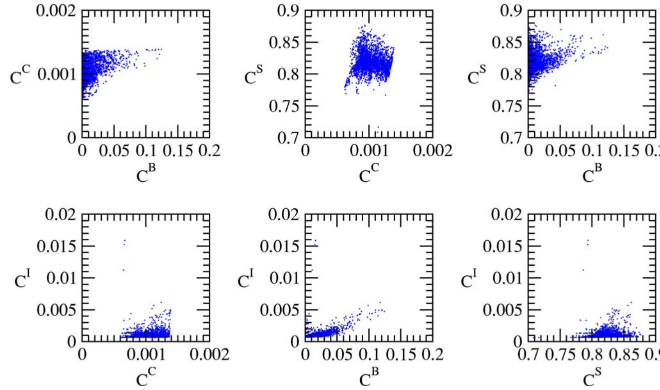

FIG. 10. Scatter plots representing all the possible correlations between C^C, C^B, C^S, and C^I. The results reported are for the city of Ahmedabad as in the preceding figure.

method, in which the distance between two clusters is defined as the longest distance from any member of one cluster to any member of the other cluster. The results obtained with single and average linkage methods are similar. The iterative pairing of cities in the dendrogram seems to capture some basic classes of urban patterns, this is the case of the early association of Barcelona and Washington or New York and Savannah, all grid-iron planned cities as well as that of Bologna, Wien, and Paris, all mostly medieval organic patterns. Brasilia, Walnut Creek, and New Delhi, to this respect, share a planned, large scale modern formation. Venice is the last association, which tells of the unique mix of fine grained pattern and natural constraints that have shaped the historical structure of the city. Choosing a maximum distance equal to 0.15 for two cities to belong to the same cluster, we find a first cluster (in red) from Barcelona to Los Angeles including medieval organic patterns and most grid-iron planned cities; a second cluster (in cyan) from Ahmedabad to Seoul including self-organized cities; a third cluster (in blue) made up by New York and Savannah, both grid-iron, but different from cities of the first cluster for peculiarities in the geometric patterns; a fourth cluster (in green) from Brasilia to New Delhi, including cities with a sizeable number of cul-de-sacs and a large scale modernist formation; a fifth cluster (in grey) constituted only by Venice, atypical for its strong natural constraints.

C. Correlations between centrality measures

Some interesting information can be extracted from the correlation between the various centrality indices. In Figs. 9 and 10 we report the scatter plots obtained for Ahmedabad, the city with the largest number of nodes, namely $N=2870$. Each point of the scatter plot represents a node of the graph. Figure 9 shows the correlation between the four measures of centrality used in the MCA, and the degree centrality, C^D (reported on the x axis). The five vertical lines corresponds to nodes, respectively, with degree k equal to 1, 2, 3, 4, and 5. Notice that in Ahmedabad there are 142 nodes with $k=1$, 11 nodes with $k=2$, 2270 with $k=3$, 435 with $k=4$, and 12 with $k=5$. The results indicate a small positive correlation between degree and betweenness, and between degree and information. Conversely, closeness and straightness are not correlated with the degree of a node. Similar results have been found for the remaining cities. In Fig. 10 we report the six scatter plots representing the correlation between all couples of indices used in the MCA. The straightness shows no significative correlation with the other three measures, in particular with closeness and betweenness. A weak positive correlation has been found between closeness and betweenness. This is true in particular for the nodes with high centrality. As already indicated in Figs. 4 and 5, information and betweenness are strongly positively correlated. A common feature of the information centrality is that the nodes with the highest value of C^I have instead a small centrality with respect to C^C, C^B, and C^S. This is a characteristic also found in the other cities.

The correlation between different centrality measures has been numerically quantified, for each of the 18 1-square-mile samples, by calculating the Pearson correlation coefficient for each couple of indices. In particular, we have investigated clustering analyses based on such correlations. For example by using the six Pearson correlation coefficients between the couples shown in Fig. 10, one obtains a hierarchical tree whose cut into five different classes gives the following clustering: Ahmenabad and Cairo in cluster 1; Bologna, Brasilia, and Savannah in cluster 2; Barcelona, Los Angeles, New Delhi, Paris, Washington, and Walnut Creek in cluster 3; London, New York, Richmond, Seoul, San Francisco, and Vienna in cluster 4; and finally Venice in cluster 5.

V. CONCLUSIONS

Analysis performed on undirected, valued, primal graphs has shown that C^C, C^B, C^S, C^I, consistently capture different natures of centrality. Despite the striking differences in terms of historical, cultural, economic, climatic, and geographic characters of selected cases, C^C, C^B, and C^S, show always the same kind of distribution. C^I, instead, is differently distributed in planned and self-organized cities, exponential for planned cities and power law for self-organized ones. The inequality of centrality indices distribution over the "population" of nodes has been investigated: A certain level of structural similarities across cities are well captured through the cluster analysis operated on the Gini coefficient. The multiple centrality assessment method, hereby presented opens up to the in depth investigation of the correlation between the structural properties of the system, and the relevant dynamics on the system,[45] like pedestrian and/or vehicular flows and other collective behaviors,[46,47] retail commerce vitality, land-use separation or urban crime, all information traditionally associated to primal graphs. We expect that some of these factors are more strictly correlated to some centrality indices than to others, thus giving informed indications on the actions that can be performed in order to increase the desired factors, as economic development, and to hinder the undesired ones, as crime rate.

An example of the possible professional applications of the method and of its relevance in the context of a problem of urban design can be found in Ref. 48. There, the MCA has been used in the problem setting phase of the program of renovation and revitalization of the open spaces of the University Campus "Area of the Sciences" in Parma, northern Italy. The method has been implemented in order to understand why the existent networks of open spaces and pedestrian paths in the Campus are scarcely experienced by students as well as faculty and staff members, and appear so poorly integrated with the life on Campus. MCA has also given a relevant contribution to the comparative evaluation of two proposed scenarios, leading to the identification of one final solution of urban design.

[1]R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2]S. H. Strogatz, Nature (London) **410**, 268 (2001).
[3]R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004).
[4]S. Wasserman and K. Faust, *Social Networks Analysis* (Cambridge University Press, Cambridge, 1994).
[5]A. Bavelas, Hum. Organ **7**, 16 (1948).
[6]G. A. Wilson, *Complex Spatial Systems: The Modeling Foundations of Urban and Regional Analysis* (Prentice-Hall, Upper Saddle River, NJ, 2000).
[7]B. Hillier and J. Hanson, *The Social Logic of Space* (Cambridge University Press, Cambridge, UK, 1984).
[8]B. Hillier, *Space is the Machine: A Configurational Theory of Architecture* (Cambridge University Press, Cambridge, UK, 1996).
[9]B. Jiang and C. Claramunt, Environ. Plan. B: Plan. Des. **31**, 15 (2004).
[10]M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).
[11]S. Porta, P. Crucitti, and V. Latora, cond-mat/0411241 (unpublished).
[12]A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[13]M. Salheen and L. Forsyth, Urb. Des. Int. 6, 93 (2001).
[14]M. T. Gastner and M. E. J. Newman, cond-mat/0407680 (unpublished).
[15]R. Larson, *Urban Operations Research* (Prentice-Hall, Upper Saddle River, NJ, 1981).
[16]F. R. Pitts, The Professional Geographer 17, 15 (1965); Soc. Networks **1**, 285 (1979).
[17]M. Batty and P. Longley, *Fractal Cities* (Academic, London, UK, 1996).
[18]C. Alexander, *The Nature of Order* (Oxford University Press, New York, 1998).
[19]N. A. Salingaros, J. Urban Design 3, 53 (1998).
[20]B. Jiang and C. Claramunt, Trans. GIS 6, 151 (2002).
[21]J. A. F. Teklenburg, H. J. P. Timmermans, and A. F. van Wagenberg, Environ. Plan. B: Plan. Des. **20**, 347 (1993).
[22]A. Penn and A. Turner, 2003, Space layout affects search efficiency for agents with vision, Proceedings 3rd International Space Syntax Symposium, London, UK, 2003.
[23]M. Batty Working Papers No. 75 and No. 80, Center for Advanced Spatial Analysis, University College London, 2004 http://www.casa.ucl.ac.uk/publications
[24]C. Ratti, Environ. Plan. B: Plan. Des. **31**, 513 (2004).
[25]B. Hillier and A. Penn, Environ. Plan. B: Plan. Des. **31**, 513 (2004).
[26]S. Porta, P. Crucitti, and V. Latora, Environ. Plan. B: Plan. Des. (to be published).
[27]P. Crucitti, V. Latora, and S. Porta, physics/0504163 (unpublished).
[28]J. Nieminen, Scand. J. Psychol. **15**, 322 (1974).
[29]L. C. Freeman, Soc. Networks **1**, 215 (1979).
[30]G. Sabidussi, Psychometrika **31**, 581 (1966).
[31]V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001).
[32]I. Vragovic, E. Louis, and A. Diaz-Guilera, Phys. Rev. E **71**, 036122 (2005).
[33]V. Latora and M. Marchiori, cond-mat/0402050 (unpublished).
[34]V. Latora and M. Marchiori, Phys. Rev. E **71**, 015103(R) (2005).
[35]A. Jacobs, *Great Streets* (MIT Press, Boston, MA, 1993).
[36]N. Dalton, J. Peponis, and R. Dalton, To tame a tiger one has to know its nature: extending weighted angular integration analysis to the description of GIS road-centerline data for large scale urban analysis, Proceedings of the 4th International Space Syntax Symposium, London, UK, 2003.
[37]A. Cardillo, S. Scellato, V. Latora, and S. Porta, physics/0510162 (unpublished).
[38]S. Scellato, A. Cardillo, V. Latora, and S. Porta, physics/0511063 (unpublished).
[39]Website at http://www.ct.infn.it/latora/1sqml.html.
[40]L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[41]P. Crucitti, V. Latora, and S. Porta (unpublished).
[42]H. A. Makse, S. Havlin, and H. E. Stanley, Nature (London) **377**, 608 (1995).
[43]K. I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[44]C. Dagum, Econ. Appl. 33, 327 (1980).
[45]B. Kuipers, D. Tecuci, and B. Stankiewicz, Environ. Behav. **35**, 81 (2003).
[46]R. H. Turner and L. M. Killian, *Collective Behavior*, 3rd ed. (Prentice Hall, Englewood Cliffs, NJ, 1987).
[47]D. Helbing, I. J. Farkas, and T. Vicsek, Nature (London) **407**, 487 (2000).
[48]S. Porta, P. Crucitti, and V. Latora (unpublished).